LAR Video: Hierarchical Representation for Low Bit-Rate Color Image Sequence Coding

Erwan Flécher, Marie Babel, Olivier Déforges, Véronique Coat

LAR VIDEO: HIERARCHICAL REPRESENTATION FOR LOW BIT-RATE COLOR IMAGE SEQUENCE CODING

Erwan Flécher, Marie Babel, Olivier Déforges, Véronique Coat

IETR CNRS UMR 6164/Image and Remote Sensing Group, INSA of Rennes
20 avenue des buttes de coësmes, 35043 Rennes cedex, France
Email: eflecher@ens.insa-rennes.fr {mbabel, odeforge, coat}@insa-rennes.fr

ABSTRACT

LAR video is a low complexity system for low bit-rate color image sequence encoding. It aims to propose a joint solution for coding and representation of the frame content. In particular, it allows to provide a compressed description of both chromatic components and motion information at a region level without region partition encoding. Initialy proposed in the LAR coder, used principle has proved to be efficient for still color image encoding. Resulting from a hierarchical spatio-temporal segmentation, a Partition Tree (PT) is transmitted to the decoder with a controlled coding cost. Presented results show interesting performances considering both content representation and compression ratios.

1. INTRODUCTION

Defined as second generation coder, region-based approaches tend to link digital systems and human perception. This type of approach provides advanced functionalities such as scalable video coding [1], region of interest (ROI) coding, video object tracking and content manipulation [2]. Despite the benefits of region-based solutions, actual standards MPEG4-AVC and SVC [3] are always based on traditional hybrid coding scheme. Principal obstacles to the content-based system evolution are: 1- an important maturation of the block-based approach 2- an incommensurate part of the bit-rate is allocated to encode region boundaries [4] especially in low bit-rate 3- high complexity of segmentation algorithms is incompatible with real-time encoding constraint. One of the most popular content-based coding system is the SESAME coder that achieves a rate-distortion (RD) optimization on a multi-scale frame representation and explicitly encodes the resulting segmentation partition. Alternative solutions attempt to not transmit the region partition coding while only considering decoded information in the segmentation process. For example, [5] presents a "symmetric-complexity coding system" where both coder and decoder compute a motion segmentation in a video coding context. Unadapted for low bit-rate, this solution produces unacceptable region description especially upon contours.

This paper presents a content-based color image sequence coding system with high compression ratio. Principal improvement of this coder called "LAR video" relies on a multi-scale representation of the frame content that is not explicitly transmitted to the decoder, contrary to [2]. A hierarchical segmentation aims to efficiently compress both color components and motion information at region levels. Restricted to the base layer encoding, the LAR video can be seen as a natural extension of the scalable still color image coder LAR [6]. The LAR (*Locally Adaptive Resolution*) is an efficient content-based solution that offerts both enhanced scalabilty at different semantic levels (pixel, block, region) and advanced functionalities such that ROI coding and lossless compression [7].

This paper is organized as follows: section 2 introduces the basic principle of the still image coder LAR that is based on a two-layer encoding scheme. Section 3 describes the segmentation process for both multi-scale representation of the image content and color encoding. An extension to the color image sequence coding that considers inter/intra prediction modes is developped in section 4. Section 5 presents some visual intermediate and decoded P-frames associated with a rate-distortion evaluation. Finaly conclusion and futur works are presented in section 6.

2. LAR FOR STILL IMAGE CODING

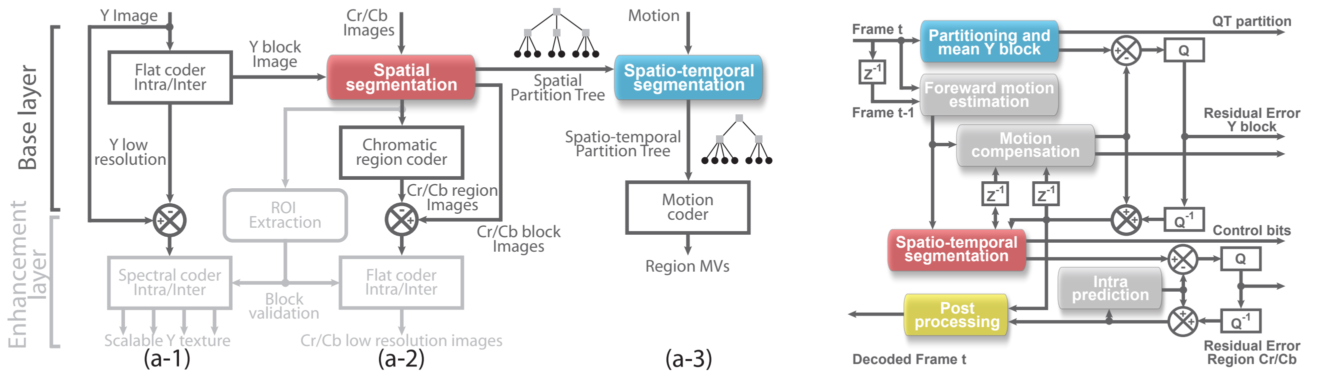
The LAR (*Locally Adaptive Resolution*) concept is based on the principle that an image can be described as the superposition of a local texture (fine details) over some low bit-rate global image information (coarse details). Thus scalability by quality levels is an inherent property of the LAR.

The LAR concept can be implemented as a two-layer coders that provides two progressivity levels at least. The base layer named *flat coder* compresses the global image information in the spatial domain whereas the local texture is encoded with a *spectral coder* that is based on a variable block-size DCT transform. Contrary to the standard scalable coder SVC [3] that encodes base and fidelity refinement layers with the same compression process, LAR coder adapts the compression technique.

2.1 Flat coder – base layer

Global image information results from a non-uniform subsampling based on a local activity analysis. On smooth luminance areas, resolution can be lowered whereas higher resolution is required on textured areas and edge positions. Thus this low complexity coder cleary aims high compression ratios. Sequential steps included in the flat coder are respectively: *partionning/modeling*, *prediction/quantization* of the model parameters and *post-processing* of the decoded image.

Partionning & modeling. Drived by a decision model, a Quadtree partitionning $P^{[N_{max},..,N_{min}]}$, where N_{max} and N_{min} are respectively the upper and lower size of square blocks, is realized. For example, standard MPEG4-AVC authorizes a partition $P^{[16,8,4]}$ for I-frame compression where the choice of the block-size is based on a rate-distortion (RD) optimization. More adapted to the human perception, an homogeneity criterion based on edge detection (using a morphological gradient) has been preferred. Moreover, a zero's order polyno-



(a) (b)

Figure 1: (a-1) Luminance Y image coder. (a-2) Region-based chromatic component (Cr/Cb) representation and compression. (a-1) Region-based motion representation and compression. (b) Simplified description of the LAR video encoder with the output stream

mial model (mean block value) is used to describe resulting block content. The model choice is in agreement with considering support size (classically $N_{max} = 16$ and $N_{min} = 2$), computational complexity and compression efficiency.

Prediction & quantization. In order to increase the compression ratio, model parameters are predicted and quantized when taking into account local image context. So a DPCM scheme based on an edge detection (using Graham predictor) that has been adapted to take advantage of the Quadtree partition is used to predict block values. Next, rate is limited thanks to an adaptive quantization considering block size.

Post-processing. Considering the model used to describe block content, a low-complexity post-processing adapted to the variable size-block representation is applied to smooth uniform zones while undamaging contours. Image resulting from this post-processing is presented in section 5.

2.2 Spectral coder – enhancement layer

To enable scalable bitstream in accordance with the required quality, an enhancement layer is added to encode error image (local texture). As both block-sizes and DC components are encoded with the flat coder, the spectral coder only works with AC components resulting from a variable block size DCT transform. More details on flat and spectral coders can be found in [7, 6].

3. HIERARCHICAL SPATIAL SEGMENTATION

Even if efficient context-based methods adapted to Quadtree-based region partition compression have been already developped [4], prohibitive partition coding cost stays one of the principal restrictions to the evolution of content-based coding solutions. Dedicated to the LAR, the segmentation proposed in [6] is an efficient adaptation of the split/merge methods that tackles coding constraints. Given that both splitting process and luminance block image encoding have been realized by the flat LAR, merging process only deals with the finest partition i.e. the Y-block image (Cr/Cb block-images are not first considered). A formal segmentation definition has been given by Horowitz and Pavlidis. Let S, $\mathcal{R}$, P be respectively a spatial domain, a region set and a predicate (associated to a homogeneity criterion). The final partition $\mathcal{R}$ of S considering predicate P is defined by:

i. $S = \bigcup_{i=1}^{|\mathcal{R}|} R_i;\ \forall i \in \{1,..,|\mathcal{R}|\}\,,\ R_i\ is\ connected$

ii. $\forall i \in \{1,..,|\mathcal{R}|\}\,,\ P(R_i) = true$

iii. $\forall (i,j) \in \{1,..,|\mathcal{R}|\}^2\ s.a.\ (R_i,R_j)\ are\ adjacent$
$\Rightarrow P(R_i \cup R_j) = false$

Originality of the merging process included in the LAR coder relies on its fusion order and its predicate P. On the one hand, segmentation allows several simultaneous merges in order to limit complexity and to provide a more compact multi-scale representation. On the other hand, a joint mean/gradient criterion weighted by region surfaces (non-symetrical distance) has been defined in order to favour regions with spatial consistency. By iteratively increasing thresholds, a hierarchical segmentation is obtained and allows to efficiently describe the image content from finest to coarse scale. Having as many indexed levels as threshold levels, indexed hierarchical segmentation can be described with a N-ary tree structure called Partition Tree PT_s^N (s: spatial; st: spatio-temporal) where N is the number of indexed levels [1, 2]. The multi-scale representation is said to be a self-extracting process (free cost) because both coder and decoder only work on Y-block image.

3.1 Non-explicit PT coding for color representation

Region-based chromatic component representation is an efficient solution that provides high compression ratio and good visual quality. To take advantage of color information, a "chromatic control" principle is defined and included in the merging process previously described. This chromatic control generates binary information for each luminance-based merging attempt to control the merging process. Moreover weighting the color information in the homogeneity criterion, a coefficient $CoefChrom$ may be considered as an ajustment cursor located between control cost and segmentation quality. Figure 1.b shows the inputs and outputs of the spatial segmentation for region-based color representation and coding. Once a level of the hierarchical segmentation has been selecting, mean chromatic values (zero's order polynomial model) are predicted, quantized and transmitted with high compression ratio.

4. REGION-BASED IMAGE SEQUENCE CODING

Proposed color image sequence coding system is based on Intra (I) and Predictive (P) frame compression where only the low bit-rate base layer is encoded. Enhancement would be considered by adding a refinement layer based on a DCT transformation. On the one hand, luminance I-frames are efficiently compressed with the flat LAR (intra prediction mode) previously described. On the other hand, inter prediction mode of the flat LAR that uses temporal redundancy between consecutive frames, has been developped to compress luminance P-frames. LAR video coder relies principaly on the multi-scale representation of the frame content that is present both at coder and decoder. This allows to efficiently encode chromatic components and to derivate a region-based representation of the motion information. The principle of the low complexity coder can be described with three sequential steps:

Step 1: Luminance block image is compressed using inter/intra prediction modes and region-based motion compensation. (Figure 1.a-1).

Step 2: Considering Y-block image and chromatic control, multi-scale spatial segmentation for region-based chromatic components representation is performed. It allows color image compression at region level and non-explicite PT_s^N encoding (Figure 1.a-2).

Step 3: Using motion feature, a hierarchical spatio-temporal segmentation is derivated of PT_s^N by split/merge process. It provides a compressed region-based description of motion between frames (Figure 1.a-3).

To summary, for each frame, the bitstream includes block sizes, quantized prediction errors of the Y-block image and region-based Cr/Cb, prediction errors of region-based motion vectors and some overhead information such that chromatic control bits and motion split/merge. (Figure 1.b) gives a simplified description of our coder with the output stream.

4.1 Luminance block image prediction & coding

Classical hybrid coders compute the Quadtree partitionning $P^{[N_{max},\ldots,N_{min}]}$ in order to decrease the DFD (Displaced Frame Difference) energy that results from the difference between current image and a motion compensated reference image. At the opposite, in the LAR video coder, partition is directly computed on the current frame in order to provide the luminance block image previously described. Inter/intra prediction modes and quantization step are then used to reduce the bit-rate of the Y-block image. P-frames are predicted with a motion-compensated Y-block image that results from a *forward motion estimation/compensation* step. This means that the previous decoded frame is projected on the current Quadtree partition using region-based motion vectors. Consequently, three types of blocks (defined in the current Quadtree partition) result from the compensation step: blocks with only one projected value, blocks with multiple projected values (*overlapped blocks*) and blocks with zero projected value (*uncovered blocks*). In two first cases, mean value is computed and is used as block prediction. In last case (i.e uncovered blocks or I-frames) blocks are predicted in intra mode with the same edge-based predictor that is used for still image encoding. Note that high prediction errors (included overlapped blocks) are classicaly located on moving object boundaries especially when region-based motion compensation is operated. This means that DFD energy is principaly concentrated on small block which are efficiently compressed with the block-size adapted quantization.

4.2 Hierarchical spatial segmentation

Once Y-block image has been compressed and transmitted to the decoder, spatial segmentation that only considers information of the current frame is processed independently for I and P-frames. This means that temporal consistency is not guaranteed. Nevertheless this is not an overriding contrary to solutions that use temporal prediction to compress the current partition considering previous decoded one. In the LAR video case, hierarchical region representation is self-extracted or transmitted with a minimal cost.

4.3 Hierarchical spatio-temporal segmentation

In section 4.1, a region-based motion compensation has been considered for P-frame encoding. Indeed, a region-based motion estimation/compensation provides better performances than classical block-based approaches, especially upon moving contours. Thus spatio-temporal segmentation aims to provide a multi-scale representation in which defined regions respect spatial and temporal homogeneities, sharing common features: gray level, color and motion. Figure 1.c gives a basic description of the spatio-temporal segmentation and motion encoding. As the Partition Tree PT_s^N is known at the decoder, the aims is to derivate a new Partition Tree PT_{st}^N based on the spatio-temporal segmentation. Sequential splitting (property *iii* of the Horowitz et al. definition) and merging (property *ii*) steps are realized in order to decompose regions if non-homogeneous motion is observed and to group regions with spatial and temporal similarities. Information associated to motion-based splitting/merging is transmitted because the decoder does not know region-based motion parameters. Note that PT_{st}^N describes an efficient research-space adapted to RD optimization because of the indexed hierarchy [1]. To describe region-based motions, a translation model has been adopted. Low complexity parameter computation and efficient parameter prediction have motivated the choice of this model instead of more sophisticated ones.

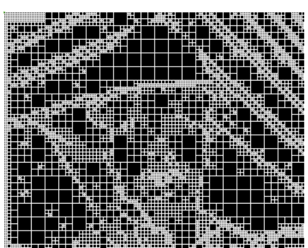
In order to take into account motion information, a fast and variable block-size motion estimator called EPZS (Enhanced Predictive Zonal Search) is initialy used to compute the forward motion vector of each region (or block) of the finest partition $P^{[N_{max},\ldots,N_{min}]}$. More precisely, motion estimation on supports with unconsistency size (classically 2×2 block-size) are not computed but result from an interpolation process. Finaly, once a level of the hierarchical spatio-temporal segmentation has been selecting, motion vectors associated to the regions are predicted and transmitted.

5. RESULT

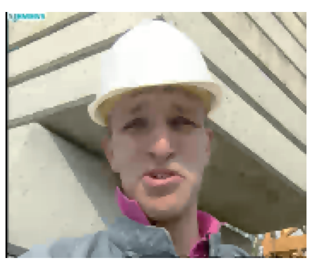
In this section, some visual and rate-distortion results are described. Figure 2 presents some intermediate images (frames 138/139/140 of Foreman_CIF @25frames/s) which show the temporal evolution of the LAR video coder. Figure 2.a is the projection of the decoded frame 138 on the current Quadtree partition (139) where dark spots describe uncovered blocks which are principally located on moving object boundaries. Figure 2.b is the decoded Y-block image (frame 139 without the post-processing). Figure 2.c is the motion field resulting from a level of the spatio-temporal segmentation (300 regions) and associated to each block of the restricted partition (139) $P^{[16,..,2[}$.

Figure 2: (a) Motion-compensated Y-block image (b) Decoded Y-block image (c) Motion field associated to the region-based representation

Resulting from the Foreman_CIF encoding with a bitrate #450kbit/s, Figure 3.b shows the decoded P-frame 50 with a PNSR #29/36/36 dB for Y and Cr/Cb components (4:2:0). The luminance is encoded with the flat LAR (inter/intra mode) and the chromatic component representation (120 regions) can be described with the region adjacency graph (RAG) presented by the Figure 3.b. Though compression results are lower than the standard MPEG4-AVC [3], the functionalities offered by this multi-scale representation with a controlled coding cost are very interesting. For example Figure 3.c show two indexed segmentation levels respectivly with 63 (left) and 359 regions (right).

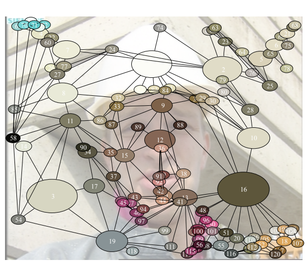


(a)



(b)

(c)



(d)

Figure 3: (a) Quadtree partition $P^{[16,..,2]}$, 0.041 bpp (b) Decoded low resolution Y-image + region-based Cr/Cb. PNSR #29/36/36 dB. (c) Two indexed levels of the spatial segmentation: left 69 regions, right: 400 regions. Chromatic control 0.015 bpp (d) RAG with 120 regions

Table 1 describes rate-distortion results. Mean rate (kbit/s) associated to the zero's order entropy and mean PSNR (dB) are compared for some color image sequences (CIF@25Hz).

	CIF (352×288)@25Hz		
Image sequences	**Rate (kbit/s)**	**PSNR Y**	**PSNR Cr/Cb**
Foreman	450	29.19	36.57/36.02
Akiyo	220	31.55	38.38/35.77
Mobile	950	21.97	29.04/31.16
Football	600	33.96	37.18/35.31
CoastGuard	650	26.27	38.18/37.8

Table 1: Rate-distortion (mean value) results for some classical color image sequences

6. CONCLUSION

We have presented a low complexity system for low bit-rate color image compression. Thanks to our non-explicit Partition Tree encoding, the proposed solution offers both compression efficiency and multi-scale content representation. This work takes place in the developpement of a global video object coder. So, futur work aims to assure temporal consistency in order to increase the compression ratio and to allow region matching/tracking between successive frames. This is primordial to consider avanced functionality such as ROI coding and video object tracking.

REFERENCES

[1] J. R. Morros and F. Marquès, "A Proposal for Dependent Optimization in Scalable Region-Based Coding Systems," in *IEEE International Conference on Image Processing, ICIP'99*, Kobe, Japan, October 1999, vol. 4, pp. 295–299.

[2] F. Marquès, P. Salembier, M. Pardas, R. Morros, I. Corset, S. Jeannin, B. Marcotegui, and F. Meyer, "A SEgmentation-based coding System Allowing Manipulation of objEcts (SESAME)," in *International Conference on Image Processing, ICIP'96*, 16-19 Sepember 1996, vol. 3, pp. 679–682.

[3] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz and M. Wien, "Joint Draft 9 of SVC Amendment," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Doc. JVT-V201*, 2007.

[4] T. Ochotta and D. Saupe, "Edge-based partition coding for fractal image compression," *The Arabian Journal for Science and Engineering, Special Issue on Fractal and Wavelet Methods*, vol. 29, no. 2C, pp. 63–83, December 2004.

[5] X. Yang and K. Ramchandran, "A low-complexity region-based video coder using backward morphological motion field segmentation," in *IEEE Transactions on Image Processing*, March 1999, vol. 8, pp. 332–345.

[6] O. Déforges, M. Babel, L. Bédat, and J. Ronsin, "Color LAR codec: a color image representation and compression scheme based on local resolution adjustment and self-extraction region representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 974–987, August 2007.

[7] M. Babel, O. Déforges, and J. Ronsin, "Interleaved S+P Pyramidal Decomposition with Refined Prediction Model," in *IEEE International Conference on Image Processing, ICIP'05*, September 2005, vol. 2, pp. 750–753.